5.0 Test Method Data and Results

5.1 Test Method Protocol

The Standard Operating Procedures (SOPs) used during the ECVAM validation studies are included in **Appendix A**. As described in **Section 2.1**, there are many similarities among the protocols for each of the *in vitro* pyrogen test methods, with very few notable differences other than the type of cells used (i.e., WB cells, PBMCs, monocytoid cell line) and the proinflammatory cytokine assayed (i.e., IL-1β or IL-6). These similarities and differences are outlined in **Table 2-1**. An internationally accepted endotoxin standard (i.e., WHO-LPS 94/580 [*E. coli* 0113:h10:K-]) was used to spike samples of saline or marketed parenteral pharmaceuticals. The same pharmaceuticals were used to create the spiked samples for all five test methods (see **Table 3-1** and **3-2**). These samples were included in a series of studies designed to determine the relevance and reliability of each of the *in vitro* pyrogen test methods.

5.2 Availability of Copies of Original Data Used to Evaluate Test Method Performance

ECVAM provided raw data from the validation studies in an electronic format (Excel® spreadsheets) that consisted of OD_{450} measurements for all replicates included in each of the validation studies.

NICEATM attempted to obtain additional *in vitro* and/or *in vivo* pyrogen test method data. A *Federal Register* (*FR*) notice (Vol. 70, No. 241, pp. 74833-74834, December 16, 2005) was published requesting original *in vitro* pyrogen test method and reference data from the currently used pyrogen test methods (i.e., RPT and/or BET). In addition, the *FR* notice was sent directly to more than 100 interested stakeholders internationally. Despite these efforts, no additional data were submitted.

5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data

Details of the statistical approaches used to evaluate the accuracy and reliability of each of the five *in vitro* test methods are included in Section 5.3 of each ECVAM BRD. Briefly, as indicated in **Section 3.2**, 10 substances (each spiked with four concentrations of endotoxin, with one concentration spiked in duplicate) were tested in each test method to evaluate accuracy, while three substances (each spiked with three concentrations of endotoxin, with one tested in duplicate) were used to evaluate test method reproducibility. Varying concentrations of endotoxin-spiked saline were tested for the analysis of intralaboratory repeatability.

The evaluation of intralaboratory repeatability included coefficient of variation (CV) analysis of the log-transformed OD_{450} measurements for the replicates of each endotoxin concentration. Boxplots were also generated to demonstrate variability among these values for each concentration. Similar analyses were conducted for the three substances used to assess intra- and inter-laboratory reproducibility.

The reproducibility analysis incorporated the decision criteria that were developed to differentiate between pyrogenic and non-pyrogenic materials (using a threshold value of 0.5

EU/mL). In all reproducibility analyses, a single run consisted of each of the substances (as described above and in **Section 3.2**) assayed in quadruplicate. Acceptability criteria for each run included a CV analysis to remove highly variable samples from the analyses. This criterion ranged from a CV<0.25 to <0.45, depending on the test method being considered. For the measurement of intralaboratory reproducibility, pair-wise comparisons between the runs were determined and the associations between runs expressed as a percentage of agreement between two individual laboratories. It should be noted that this analysis takes into account the agreement of the resulting pyrogenicity decision (i.e., pyrogenic or non-pyrogenic), but does not consider whether the decision is correct. The correlations (expressed as a percentage of agreement) between pairs of the independent runs (i.e., run 1 vs. run 2; run 1 vs. run 3; run 2 vs. run 3) were determined and the mean of these three values was calculated. Similar analyses were conducted for an assessment of interlaboratory reproducibility, in which pairwise comparisons between laboratories were determined and the associations were expressed as a percentage of agreement. This analysis included each run from each laboratory (n=3 per laboratory) and all possible interlaboratory combinations were compared. Similar to the intralaboratory analysis, this analysis takes the resulting pyrogenicity call from each run in each laboratory into consideration, but does not consider whether the call is correct. **Section 7.0** provides additional details and the resulting data from these analyses.

For the accuracy analysis, 2 x 2 contingency tables were constructed using the decision criteria defined in **Table 4-2** to assign a pyrogenicity call. Each run for each sample from each laboratory was considered independently. Accordingly, the *in vitro* call was compared to the "true status" (based on the known endotoxin spike concentration) of the sample. The resulting accuracy statistics were calculated based on the overall database for each test method. Similar to the reproducibility analyses, acceptability criteria for each run included a CV analysis to remove highly variable samples from the analyses, for which a range of CV<0.25 to CV<0.45 was used, depending on the test method being considered. **Section 6.0** provides additional details and the resulting data from these analyses.

Outliers were identified and eliminated using a two-step procedure. In the first step, replicates with an extremely large variation were identified by comparing the CV for the replicates with the extracted maximal CV (CV_{max}). If the CV for the replicates was smaller than the CV_{max}, then the data were analyzed without modification. However, if the replicates failed to pass this initial test, then the data were transformed with the natural logarithm and examined for outliers using the nonparametric Dixon's test (Dixon 1950; Barnett and Lewis 1984) or the Grubbs' test (Grubbs 1969) for normally distributed samples. If one observation was responsible for the large variation, then the observation was excluded. If the variation was due to all observations, then the entire set of replicates was excluded from further analysis. Additional information on the analytical procedure used to identify and eliminate outlier observations can be found in the materials provided by ECVAM (see **Appendix C**).

5.4 Summary of Results

Graphical representations of the repeatability and reproducibility analyses are provided in Section 5.2 of each ECVAM BRD (see **Appendix A**). The tabulated results from which the intra- and inter-laboratory reproducibility analyses and accuracy analyses can be conducted are provided in Section 5.4 of the ECVAM BRDs. The tables in that section include the test

substance name, the endotoxin spike concentration, the pyrogenicity call for each *in vitro* run, and the "true status" of each test substance.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained using coded chemicals and reported in accordance with GLP guidelines (i.e., OECD 1998; EPA 2003a, 2003b; FDA 2003). **Section 3.4** indicates that the 10 test substances and the four spike concentrations used for the accuracy evaluation were blinded to the testing laboratories. However, although the three spike concentrations were blinded to the participating laboratories for the reproducibility studies, the identity of the three test substances was not blinded.

5.6 Lot-to-Lot Consistency of Test Substances

Lot-to-lot consistency of test substances is evaluated to ensure that the same substance, with the same physicochemical properties, is used for the duration of the study. In these studies, the test substances were released from clinical lots of parenteral pharmaceuticals, which implied that they had been subjected to rigorous chemical manufacturing control analyses to verify that the compositions are consistent. However, the specific lot numbers for the test substances used in the validation study were not initially provided in the ECVAM BRDs. In response to a request for additional information, ECVAM provided this information (**Table 3-1** and **Appendix C**). In addition, the international standard for Gram-negative endotoxin, WHO-LPS 94/580 (*E. coli* O113:H10:K-), was used as the spike solution, which provides a measure of consistency for the positive control substance and the spike substance.

5.7 Availability of Data for External Audit

As described in **Section 8.4**, all records are stored and archived by the participating laboratories and are available for inspection.